20th March, 2024

STK4141/9141 - Probabilistic Graphical Models

Mandatory assignment 1 of 1

Submission deadline

Thursday 18th APRIL 2024, 14:30 in Canvas (canvas.uio.no).

Instructions

Note that you have **one attempt** to pass the assignment. This means that there are no second attempts.

You can choose between scanning handwritten notes or typing the solution directly on a computer (for instance with LaTeX). The assignment must be submitted as a single PDF file. Scanned pages must be clearly legible. The submission must contain your name, course and assignment number.

It is expected that you give a clear presentation with all necessary explanations. Remember to include all relevant plots and figures. All aids, including collaboration, are allowed, but the submission must be written by you and reflect your understanding of the subject. If we doubt that you have understood the content you have handed in, we may request that you give an oral account.

In exercises where you are asked to write a computer program, you need to hand in the code along with the rest of the assignment. It is important that the submitted program contains a trial run, so that it is easy to see the result of the code.

Application for postponed delivery

If you need to apply for a postponement of the submission deadline due to illness or other reasons, you have to contact the Student Administration at the Department of Mathematics (e-mail: studieinfo@math.uio.no) no later than the same day as the deadline.

All mandatory assignments in this course must be approved in the same semester, before you are allowed to take the final examination.

Complete guidelines about delivery of mandatory assignments:

uio.no/english/studies/admin/compulsory-activities/mn-math-mandatory.html

The assignment consists of five problems (Problem 1–5), which are the same for both STK4141 and STK9141. However, for PhD candidates who have registered for STK9141, there is one additional assignment given in Problem 6, which students in STK4141 may disregard. To pass the assignment, the presented solutions do not have to be entirely correct, but you are expected to make a genuine effort at solving each problem (which should also be reflected by the submitted report).

Problem 1. Mutual information

In this problem, we are going to prove some properties related to mutual information, which can be used to measure the strength of a statistical dependence. In particular, to measure the conditional dependence between X and Y given some variable Z, we can use conditional mutual information:

$$\mathbb{I}_P(X;Y \mid Z) = \sum_{x,y,z} P(x,y,z) \log\left(\frac{P(x \mid y,z)}{P(x \mid z)}\right).$$

(a) Prove the chain rule of mutual information:

$$\mathbb{I}_P(X;Y,Z) = \mathbb{I}_P(X;Y) + \mathbb{I}_P(X;Z \mid Y).$$

and give an intuitive explanation of what it entails.

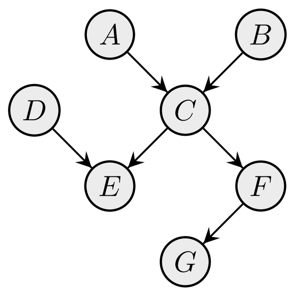
(b) Prove that mutual information is non-negative.

(c) Prove that the information Y and Z provide about X cannot be less than the information Y provides on its own, that is, prove the inequality:

$$\mathbb{I}_P(X;Y) \leq \mathbb{I}_P(X;Y,Z).$$

Problem 2. Minimal I-Map under Marginalization

In some applications we might want to marginalize out a node from a network. Consider the following Bayesian network structure, $\mathcal{G}$:



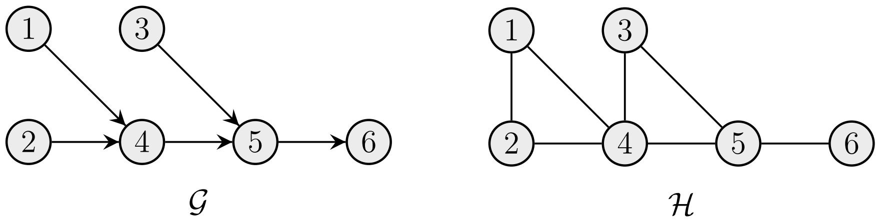
Construct a Bayesian network structure $\mathcal{G}'$ over all nodes except C such that $\mathcal{G}'$ is a minimal I-map for the marginal distribution:

$$P(A, B, D, E, F, G) = \sum_C P(A, B, C, D, E, F, G).$$

Make sure that all dependencies in $\mathcal{G}$ are maintained in $\mathcal{G}'$.

Problem 3. Markov Properties and Factorization

Let $P = P(\mathcal{X})$ denote a joint distribution over variables $\mathcal{X} = \{X_1, \ldots, X_6\}$. Further, let the following directed acyclic graph (DAG), $\mathcal{G}$, and undirected graph, $\mathcal{H}$, represent the dependence structure of a Bayesian network and Markov network, respectively, over $\mathcal{X}$:



(a) Let $\mathcal{G}$ be an I-map for P. Is $\mathcal{H}$ an I-map for P? Motivate your answer.

(b) Let $\mathcal{H}$ be a perfect map for P. Is there a DAG $\mathcal{G}'$ which is a perfect map for P? Motivate your answer.

(c) Let P be a distributions that factorizes as over $\mathcal{G}$. Show that P also factorizes over $\mathcal{H}$ according to

$$P(X_1, \ldots, X_6) = \frac{\prod_{\boldsymbol{C}_i \in \mathcal{C}} P(\boldsymbol{C}_i)}{\prod_{\boldsymbol{S}_j \in \mathcal{S}} P(\boldsymbol{S}_j)},$$

where $\mathcal{C}$ is the set of maximal cliques and $\mathcal{S}$ is the set of separators (or sepsets), which in this case is the set of non-empty intersections for each pair of maximal cliques $\boldsymbol{C}_i, \boldsymbol{C}_j \in \mathcal{C}$ where $i \neq j$.

Problem 4. Log-Linear Models

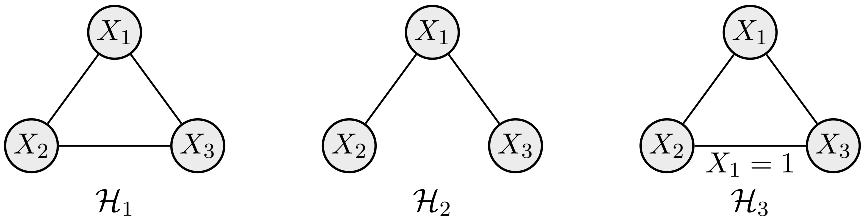
A common way to parameterize a Markov network with a positive distribution P is through a log-linear model. Assume that we have three binary variables:

$$\mathcal{X} = \{X_1, X_2, X_3\}, \quad \text{where } \mathrm{Val}(X_i) = \{0, 1\} \text{ for } i = 1, 2, 3.$$

Now, assuming a full graph $\mathcal{H}_1$ (no independence restrictions) we specify the associated distribution by

$$P_{\mathcal{H}_1}(x_1, x_2, x_3) = \frac{1}{Z}\exp(\phi_1 x_1 + \phi_2 x_2 + \phi_3 x_3 + \phi_{12} x_1 x_2 + \phi_{13} x_1 x_3 + \phi_{23} x_2 x_3 + \phi_{123} x_1 x_2 x_3),$$

where each $\phi \in \mathbb{R}$.



(a) Assume the model structure $\mathcal{H}_2$, which encodes that $(X_2 \perp X_3 \mid X_1)$. Under the given assumptions, show that the conditional independence restriction is equivalent to the following restrictions on the joint distribution:

$$\frac{P(X_1 = x_1, X_2 = 0, X_3 = 1)}{P(X_1 = x_1, X_2 = 0, X_3 = 0)} = \frac{P(X_1 = x_1, X_2 = 1, X_3 = 1)}{P(X_1 = x_1, X_2 = 1, X_3 = 0)}, \quad x_1 \in \{0, 1\}.$$

(b) What type of restrictions will $\mathcal{H}_2$ impose on the model parameters, that is, the ϕ parameters?

Context-specific independence is a generalization of conditional independence that may hold only in a specified sub-space of the conditioning variables. Thus, a conditional independence statement can be expressed as a collection of context-specific independence statements:

$$(X_2 \perp X_3 \mid X_1) \Leftrightarrow \{(X_2 \perp X_3 \mid X_1 = x_1)\}_{x_1 \in \mathrm{Val}(X_1)}.$$

(c) Assume the model structure $\mathcal{H}_3$, where the labeled edge encodes the context-specific independence $(X_2 \perp X_3 \mid X_1 = 1)$. Again, note that $(X_2 \perp X_3 \mid X_1 = 0)$ is not implied in this case. What type of restriction will $\mathcal{H}_3$ impose on the model parameters?

Problem 5. Structure Learning for Bayesian Networks

For the first part of the problem, we are going to use the data set `bn_data.csv`, which can be downloaded from the course semester page.

The data is in csv-format and it contains 1000 observations (rows) over 11 trinary variables (columns), with values $\{0, 1, 2\}$. The data has been generated by a Bayesian network, which we assume is unknown, and we want to learn about the structure of the network.

(a) Implement the maximum weight spanning tree algorithm (Alg A.2 in the PGM-book) and use it to find the maximum-likelihood tree for the given data[1]. Draw the graph.

(b) What is the BIC score of the tree found in (a)?

(c) Find a DAG that has a higher BIC score than the tree (the higher the better:). You may use functions from existing packages, e.g. the ones we have looked at in the weekly coding exercises. Draw the graph.

For the second part of the problem, we are going to use the *Congressional Voting Records data set*[2], which can be downloaded from the UCI Machine Learning Repository. The data set includes the political party affiliation and voting records for 16 key votes of congress members from the U.S. House of Representatives. A formatted version of the data can be downloaded directly from the course semester page (`congress_data.csv`). In this data, the categories have been transformed into integers according to

$$\text{Column 1: } \{\text{Democrat, Republican}\} \rightarrow \{0, 1\},$$
$$\text{Columns 2-17: } \{\text{No, Yes, ?}\} \rightarrow \{0, 1, 2\},$$

thus following the same format as the first data set.

(d) Use your implemented function from (a) to find the maximum likelihood tree for the congressional voting records data and draw the graph. Based on the connectivity in the graph, which questions appear to be of particular importance in terms of explaining the observed voting patterns?

Note: Remember to also include your code used to produce the results.

[1] The optimal tree should have a log-maximum-likelihood score of -7659.42.

[2] Schlimmer, J. C. (1987). Concept acquisition through representational adjustment. Doctoral dissertation, Department of Information and Computer Science, University of California, Irvine, CA.

Problem 6. Presentation (Only for STK9141)

As part of the mandatory assignment, each PhD candidate is expected to give a 20-minute oral presentation on a topic related to probabilistic graphical models. The presentation will be given during a lecture after the submission deadline. As part of the report, write down a title for the presentation along with a brief abstract (a few sentences) describing the planned presentation. Feel free to confer with the lecturer if you are uncertain about the topic.

Good luck!